

DECODING MULTIPLE HMM SETS USING A SINGLE SENTENCE GRAMMAR

this claims benefit of 60/2219207B1/00

Field of Invention

[0001] This invention relates to speech recognition and more particularly to a speech recognition search method.

Background of Invention

[0002] Speech recognition devices are typically deployed in different acoustic environments. An acoustic environment refers to a stationary condition in which the speech is produced. For instance, speech signal can be produced by male speakers, female speakers, in office environment, in noisy environment.

[0003] A common way of dealing with multiple environment speech recognition is to train a sets of models such as Hidden Markov Models (HMM) for each environment. Each set of HMMs will have the same number of models representing the same sounds or words spoken in the environment corresponding to the HMM set. Typically, a speech recognizer utilizes a grammar network which specifies the sequence of HMMs that correspond to the particular speech sounds and words making up the allowable sentences. In order to handle the sets of HMMs for each environment, current art technology provides the speech recognizer with a large grammar network which contains a grammar sub-network for each HMM set according to each of the environments. These sub-networks enable the use of each HMM sets within the recognizer. Since the HMM sequences corresponding to sentences allowed by the grammar network generally do not change with environment each grammar sub-network has the same structure. For example, there would be a pronunciation set or network of HMMs (grammars) for male speakers and a set of HMMs for female speakers because the sounds or models for a male speaker are different from a female speaker. At the recognition phase, HMMs of all environments are decoded and the recognition result of the environment giving the maximum likelihood is considered as final results. Such a practice is very efficient in recognition performance. For example, if male/female separate models are not used, with the same amount of HMM parameters, the Word Error Rate (WER) will typically increase 70%.